



OIKE

RAW SEQUENCE LISTING

DATE: 05/31/2002

PATENT APPLICATION: US/10/082,671

TIME: 12:15:56

Input Set : A:\Xen-001.app

Output Set: N:\CRF3\05312002\J082671.raw

p6

3 <110> APPLICANT: DAHIYAT, BASSIL
 4 LI, MIN
 6 <120> TITLE OF INVENTION: USE OF NUCLEIC ACID LIBRARIES TO CREATE TOXICOLOGICAL
 7 PROFILES
 9 <130> FILE REFERENCE: XEN/001
 11 <140> CURRENT APPLICATION NUMBER: 10/082,671
 12 <141> CURRENT FILING DATE: 2002-05-17
 14 <150> PRIOR APPLICATION NUMBER: 60/270,781
 15 <151> PRIOR FILING DATE: 2001-02-22
 17 <160> NUMBER OF SEQ ID NOS: 58
 19 <170> SOFTWARE: PatentIn Ver. 2.1
 21 <210> SEQ ID NO: 1
 22 <211> LENGTH: 9
 23 <212> TYPE: PRT
 24 <213> ORGANISM: Artificial Sequence
 26 <220> FEATURE:
 27 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 28 peptide
 30 <220> FEATURE:
 31 <221> NAME/KEY: MOD_RES
 32 <222> LOCATION: (1)..(3)
 33 <223> OTHER INFORMATION: Variable amino acid
 35 <220> FEATURE:
 36 <221> NAME/KEY: MOD_RES
 37 <222> LOCATION: (6)
 38 <223> OTHER INFORMATION: Variable amino acid
 40 <220> FEATURE:
 41 <221> NAME/KEY: MOD_RES
 42 <222> LOCATION: (8)..(9)
 43 <223> OTHER INFORMATION: Variable amino acid
 45 <400> SEQUENCE: 1
 46 Xaa Xaa Xaa Pro Pro Xaa Pro Xaa Xaa
 47 1 5
 50 <210> SEQ ID NO: 2
 51 <211> LENGTH: 20
 52 <212> TYPE: PRT
 53 <213> ORGANISM: Artificial Sequence
 55 <220> FEATURE:
 56 <223> OTHER INFORMATION: Description of Artificial Sequence: Consensus
 57 sequence for SH-3 domain binding protein
 59 <220> FEATURE:
 60 <221> NAME/KEY: MOD_RES
 61 <222> LOCATION: (3)..(7)

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/082,671

DATE: 05/31/2002

TIME: 12:15:56

Input Set : A:\Xen-001.app

Output Set: N:\CRF3\05312002\J082671.raw

```

62 <223> OTHER INFORMATION: Unknown amino acid
64 <220> FEATURE:
65 <221> NAME/KEY: MOD_RES
66 <222> LOCATION: (13)
67 <223> OTHER INFORMATION: Val, Ala, Gly, Leu, Pro or Arg
69 <220> FEATURE:
70 <221> NAME/KEY: MOD_RES
71 <222> LOCATION: (15)..(16)
72 <223> OTHER INFORMATION: Val, Ala, Gly, Leu, Pro or Arg
74 <400> SEQUENCE: 2
W-> 75 Met Gly Xaa Xaa Xaa Xaa Xaa Arg Pro Leu Pro Pro Xaa Pro Xaa Xaa
76      1              5              10              15
78 Gly Gly Pro Pro
79      20
82 <210> SEQ ID NO: 3
83 <211> LENGTH: 63
84 <212> TYPE: DNA
85 <213> ORGANISM: Artificial Sequence
87 <220> FEATURE:
88 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
89      oligonucleotide consensus sequence for SH-3 domain
90      binding protein
92 <220> FEATURE:
93 <221> NAME/KEY: modified_base
94 <222> LOCATION: (7)..(8)
95 <223> OTHER INFORMATION: a, c, t or g
97 <220> FEATURE:
98 <221> NAME/KEY: modified_base
99 <222> LOCATION: (10)..(11)
100 <223> OTHER INFORMATION: a, c, t or g
102 <220> FEATURE:
103 <221> NAME/KEY: modified_base
104 <222> LOCATION: (13)..(14)
105 <223> OTHER INFORMATION: a, c, t or g
107 <220> FEATURE:
108 <221> NAME/KEY: modified_base
109 <222> LOCATION: (16)..(17)
110 <223> OTHER INFORMATION: a, c, t or g
112 <220> FEATURE:
113 <221> NAME/KEY: modified_base
114 <222> LOCATION: (19)..(20)
115 <223> OTHER INFORMATION: a, c, t or g
117 <400> SEQUENCE: 3
WY-> 118 atgggcnnkn nknnknknkn kagacctgtg cctccasbkg ggsbkbsbkkg aggccacct 60
119 taa 63
122 <210> SEQ ID NO: 4
123 <211> LENGTH: 4
124 <212> TYPE: PRT
125 <213> ORGANISM: Artificial Sequence

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/082,671

DATE: 05/31/2002

TIME: 12:15:56

Input Set : A:\Xen-001.app

Output Set: N:\CRF3\05312002\J082671.raw

```

127 <220> FEATURE:
128 <223> OTHER INFORMATION: Description of Artificial Sequence: Linker
129     consensus sequence
131 <400> SEQUENCE: 4
132 Gly Gly Gly Ser
133     1
136 <210> SEQ ID NO: 5
137 <211> LENGTH: 69
138 <212> TYPE: PRT
139 <213> ORGANISM: Artificial Sequence
141 <220> FEATURE:
142 <223> OTHER INFORMATION: Description of Artificial Sequence: Minibody
143     presentation structure
145 <400> SEQUENCE: 5
146 Met Gly Arg Asn Ser Gln Ala Thr Ser Gly Phe Thr Phe Ser His Phe
147     1             5             10             15
149 Tyr Met Glu Trp Val Arg Gly Gly Glu Tyr Ile Ala Ala Ser Arg His
150             20             25             30
152 Lys His Asn Lys Tyr Thr Thr Glu Tyr Ser Ala Ser Val Lys Gly Arg
153             35             40             45
155 Tyr Ile Val Ser Arg Asp Thr Ser Gln Ser Ile Leu Tyr Leu Gln Lys
156     50             55             60
158 Lys Lys Gly Pro Pro
159     65
162 <210> SEQ ID NO: 6
163 <211> LENGTH: 82
164 <212> TYPE: DNA
165 <213> ORGANISM: Artificial Sequence
167 <220> FEATURE:
168 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
169     oligonucleotide
171 <400> SEQUENCE: 6
172 aggaaccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg 60
173 ccgcccgggc aaagcccggg cg                                     82
176 <210> SEQ ID NO: 7
177 <211> LENGTH: 10
178 <212> TYPE: PRT
179 <213> ORGANISM: Artificial Sequence
181 <220> FEATURE:
182 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
183     linker
185 <400> SEQUENCE: 7
186 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
187     1             5             10
190 <210> SEQ ID NO: 8
191 <211> LENGTH: 1866
192 <212> TYPE: DNA
193 <213> ORGANISM: adeno-associated virus 2
195 <400> SEQUENCE: 8

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/082,671

DATE: 05/31/2002

TIME: 12:15:56

Input Set : A:\Xen-001.app

Output Set: N:\CRF3\05312002\J082671.raw

```

196 atgcgcgggt tttacgagat tgtgattaag gtccccagcg accttgacga gcatctgccc 60
197 ggcattttctg acagctttgt gaactgggtg gccgagaagg aatgggagtt gccgccagat 120
198 tctgacatgg atctgaatct gattgagcag gcacccctga ccgtggccga gaagctgcag 180
199 cgcgactttc tgacggaatg gcgcggtgtg agtaaggccc cggaggccct tttctttgtg 240
200 caatttgaga agggagagag ctacttccac atgcacgtgc tcgtggaaac caccggggtg 300
201 aaatccatgg ttttgggacg tttcctgagt cagattcgcg aaaaactgat tcagagaatt 360
202 taccgcggga tcgagccgac tttgccaaac tggttcgcg tcacaaagac cagaaatggc 420
203 gccggaggcg ggaacaaggt ggtggatgag tgctacatcc ccaattactt gctcccaaaa 480
204 acccagcctg agctccagtg ggcgtggact aatatggaac agtatttaag cgctgtttg 540
205 aatctcacgg agcgtaaacg gttggtggcg cagcatctga cgacgtgtc gcagacgcag 600
206 gaycagaaca aagagaatca gaatcccaat tctgatgcgc cgtgatcag atcaaaaact 660
207 tcagccaggt acatggagct ggtcgggttg ctctgggaca aqgggattac ctccggagaag 720
208 cagtggatcc agggagacca ggctcctaac atctccttca atgcggcctc caactcgcg 780
209 tcccaaatca aggtgcctt ggacaatgcg ggaaagatta tgagcctgac taaaaccgcc 840
210 cccgactacc tgggtggcca gcagcccggtg gaggacattt ccagcaatcg gatttataaa 900
211 attttggaaac taaacgggta cgatcccaaa tatgcggctt ccgtctttct gggatgggcc 960
212 acgaaaaagt tcggcaagag gaacaccatc tggctgtttg ggctgcaac taccgggaag 1020
213 accaaccatcg cggaggccat agcccacact gtgcccttct acgggtgcgt aaactggacc 1080
214 aatgagaact ttccttcaa cgaactgtgc gacaagatgg tgatctggtg ggaggagggg 1140
215 aagatgaccg ccaaggctgt ggagtcggcc aaagccattc tcggagggaag caagggtgcg 1200
216 gtggaccaga aatgcaagtc ctcgcccag atagaccgga ctcccgatg cgtcacctcc 1260
217 aacaccaaca tgtgcgccgt gattgacggg aactcaacga ccttcgaaca ccagcagccg 1320
218 ttgcaagacc ggatgttcaa atttgaactc acccgccgtc tggatcatga ctttgggaag 1380
219 gtcaccaagc aggaagtcaa agacttttcc cgttgggcaa aggatcacgt ggttgagggtg 1440
220 gagcatgaat tctacgtcaa aaaggtgga gccaaagaaa gaccgcccc cagtgcgca 1500
221 gatataagtg agcccaaacg ggtgcgcgag tcagttgcgc agccatcgac gtcagacgcg 1560
222 gaagcttcga tcaactacgc agacaggtac caaaacaaat gttctcgtca cgtgggcatg 1620
223 aatctgatgc tgtttccctg cagacaatgc gagagaatga atcagaattc aaatatctgc 1680
224 ttcactcacg gacagaaaga ctgtttagag tgctttcccg tgtcagaatc tcaaccggtt 1740
225 tctgtcgtca aaaaggcgta tcagaaactg tgctacattc atcatatcat gggaaagggtg 1800
226 ccagacgctt gcaactgcctg cgatctggtc aatgtggatt tggatgactg catctttgaa 1860
227 caataa
230 <210> SEQ ID NO: 9
231 <211> LENGTH: 621
232 <212> TYPE: PRT
233 <213> ORGANISM: adeno-associated virus 2
235 <400> SEQUENCE: 9
236 Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
237 1 5 10 15
239 Gly His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
240 20 25 30
242 Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
243 35 40 45
245 Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
246 50 55 60
248 Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
249 65 70 75 80
251 Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
252 85 90 95

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/082,671

DATE: 05/31/2002

TIME: 12:15:57

Input Set : A:\Xen-001.app

Output Set: N:\CRF3\05312002\J082671.raw

```

254 Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
255          100          105          110
257 Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
258          115          120          125
260 Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
261          130          135          140
263 Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
264 145          150          155          160
266 Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
267          165          170          175
269 Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
270          180          185          190
272 Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
273          195          200          205
275 Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
276          210          215          220
278 Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
279 225          230          235          240
281 Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
282          245          250          255
284 Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
285          260          265          270
287 Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
288          275          280          285
290 Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
291          290          295          300
293 Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
294 305          310          315          320
296 Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
297          325          330          335
299 Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
300          340          345          350
302 Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
303          355          360          365
305 Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
306          370          375          380
308 Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
309 385          390          395          400
311 Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
312          405          410          415
314 Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
315          420          425          430
317 Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
318          435          440          445
320 Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
321          450          455          460
323 Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
324 465          470          475          480
326 Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/082,671

DATE: 05/31/2002
TIME: 12:15:58

Input Set : A:\Xen-001.app
Output Set: N:\CRF3\05312002\J082671.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; Xaa Pos. 1,2,3,6,8,9
Seq#:2; Xaa Pos. 3,4,5,6,7,13,15,16
Seq#:3; N Pos. 7,8,10,11,13,14,16,17,19,20
Seq#:58; Xaa Pos. 3,4,5,6